

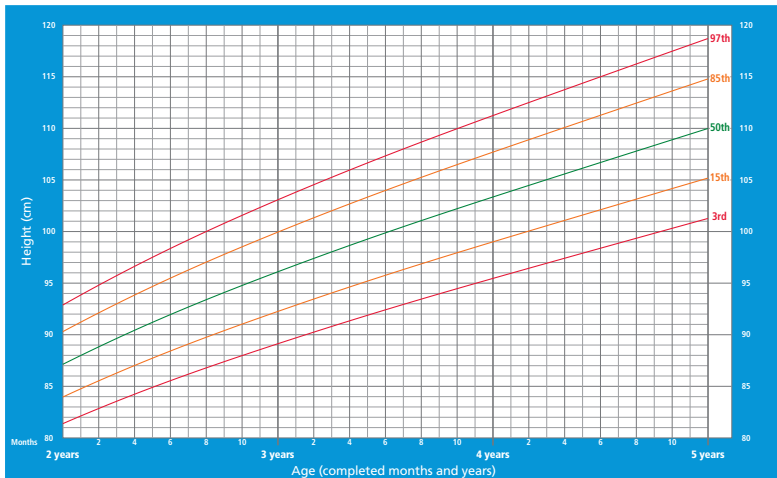
Kap 11: Enkel lineær regresjon og korrelasjon

I dag:

- Motivasjon /eksempel (Kap 11.1)
- Oppsett enkel lineær regresjon (Kap 11.2)
- Estimering: Minste kvadraters metode (Kap 11.3)
- Egenskaper for estimatorar (Kap 11.4)

Height-for-age BOYS

2 to 5 years (percentiles)

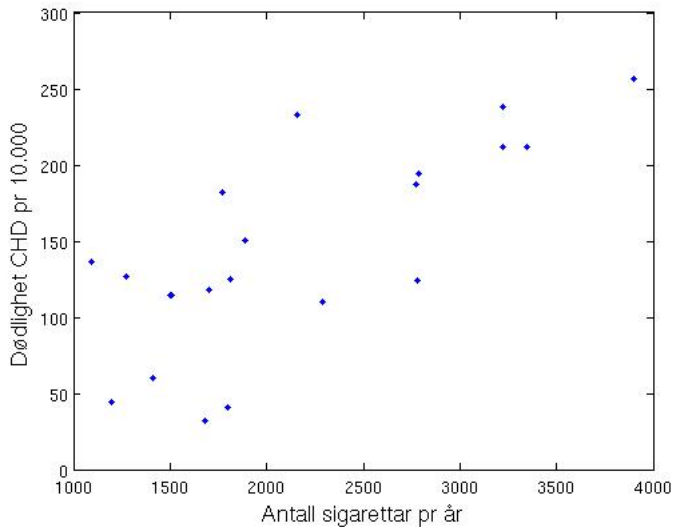


WHO Child Growth Standards

on 10.2.16 By late 1971, all cigarette packs had to be labeled with the word "Warning: The Surgeon General Has Determined That Smoking Is Dangerous to Your Health." The case against smoking rested heavily on statistical, rather than laboratory, evidence. Extensive surveys of smokers and nonsmokers had revealed the former to have much higher risks of dying from a variety of causes, including heart disease. The table below shows (1) the annual cigarette consumption and (2) the mortality rate due to coronary heart disease (CHD) for 21 countries (109).

Year	Country	Cigarette Consumption per Adult per Year	CHD Mortality per 100,000 (Ages 35-64)
1962	United States	3900	256.9
1962	Canada	3350	211.6
1962	Australia	3220	238.1
1962	New Zealand	3220	211.8
1963	United Kingdom	2790	194.1
1962	Switzerland	2780	124.5
1962	Ireland	2770	187.3
1962	Iceland	2290	110.5
1962	Finland	2160	233.1
1963	West Germany	1890	150.3
1962	Netherlands	1810	124.7
1962	Greece	1800	41.2
1962	Austria	1770	182.1
1962	Belgium	1700	118.1
1962	Mexico	1680	31.9
1963	Italy	1510	114.3
1961	Denmark	1500	144.9
1962	France	1410	59.7
1962	Sweden	1270	126.9
1961	Spain	1200	43.9

Sigarett - hjertestans



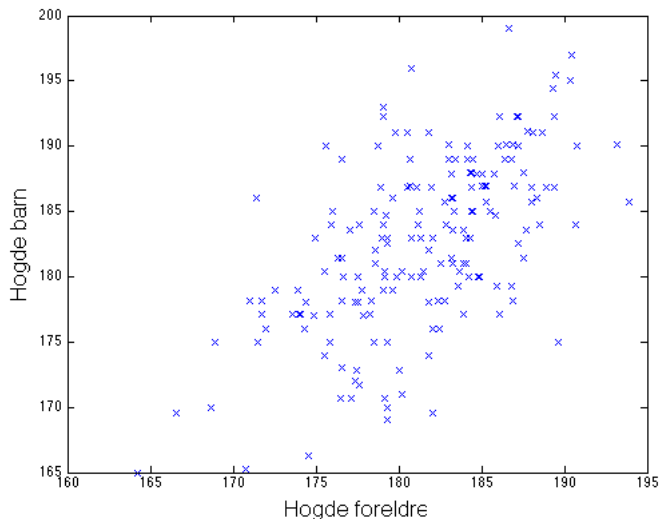
Høgde foreldre og barn, Galton

TABLE 10.3.1 NUMBER OF ADULT CHILDREN OF VARIOUS STATURES BORN OF 205 MID-PARENTS OF VARIOUS STATURES (ALL FEMALE HEIGHTS HAVE BEEN MULTIPLIED BY 1.08)

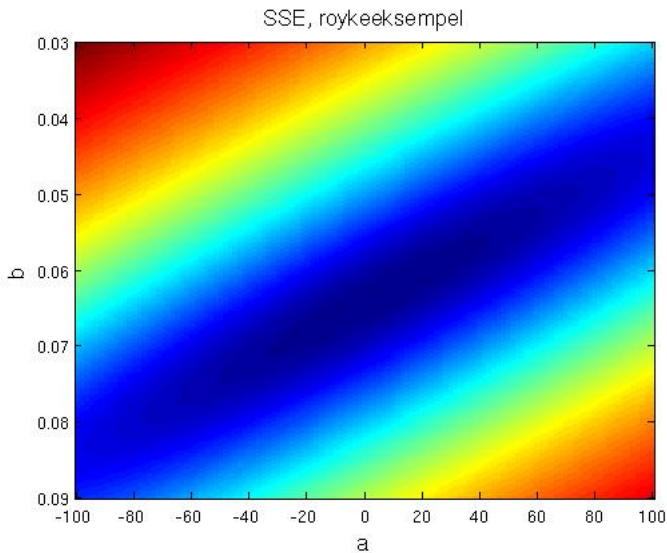
Height of the Mid-Parents (inches)	Heights of the Adult Children														Total Number of:		Medians or Values of M
	Below	62.2	63.2	64.2	65.2	66.2	67.2	68.2	69.2	70.2	71.2	72.2	73.2	Above	Adult Children	Mid-Parents	
Above 72.5	—	—	—	—	—	—	—	—	—	—	—	1	3	—	4	5	
72.5	—	—	—	—	—	—	—	1	2	1	2	7	2	4	19	6	72.2
71.5	—	—	—	—	1	3	4	3	5	10	4	9	2	2	43	11	69.9
70.5	1	—	1	—	1	1	3	12	18	14	7	4	3	3	68	22	69.5
69.5	—	—	1	16	4	17	27	20	33	25	20	11	4	5	183	41	68.9
68.5	1	—	7	11	16	25	31	34	48	21	18	4	3	—	219	49	68.2
67.5	—	3	5	14	15	36	38	28	38	19	11	4	—	—	211	33	67.6
66.5	—	3	3	5	2	17	17	14	13	4	—	—	—	—	78	20	67.2
65.5	1	—	9	5	7	11	11	7	7	5	2	1	—	—	66	12	66.7
64.5	1	1	4	4	1	5	5	—	2	—	—	—	—	—	23	5	65.8
Below 64.5	1	—	2	4	1	2	2	1	1	—	—	—	—	—	14	1	
Totals	5	7	32	59	48	117	138	120	167	99	64	41	17	14	928	205	
Medians	—	—	66.3	67.8	67.9	67.7	67.9	68.3	68.5	69.0	69.0	70.0					

Høgde foreldre-barn, våre data

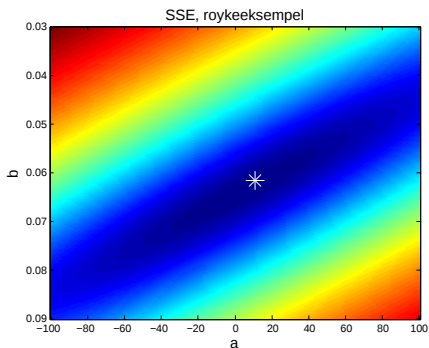
For alle kvinner: $h_{korr} = 1.08 \cdot h$.



SSE(a,b), sigarett-hjertestans

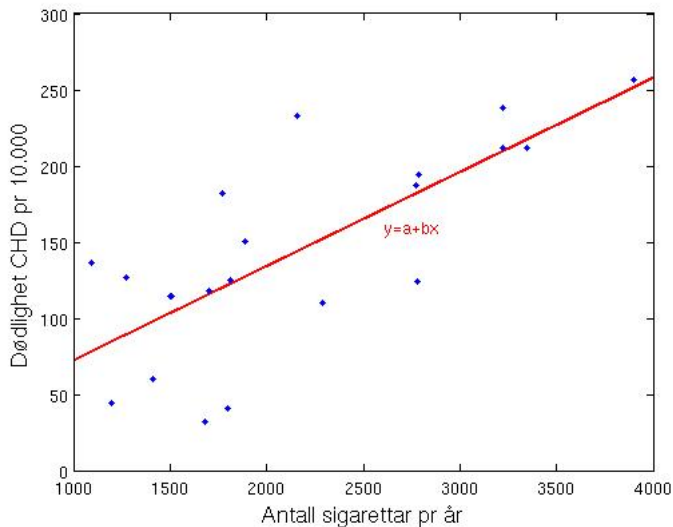


SSE(a,b), minimum (*) sigarett-hjertestans



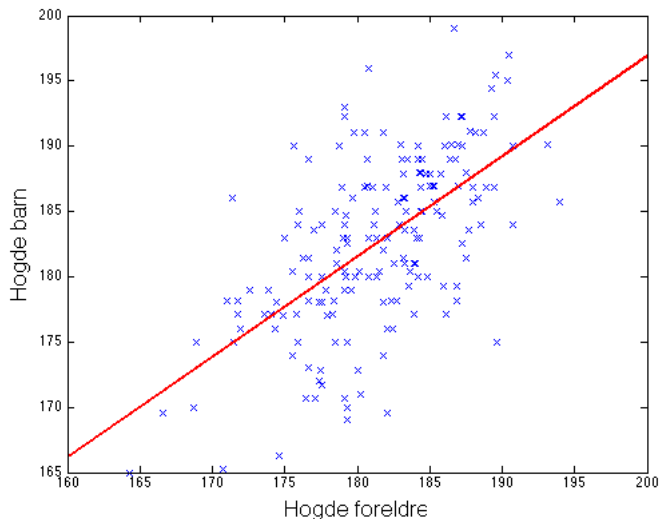
Sigarett - hjertestans

Estimat: $a = 11.41$ og $b = 0.0616$



Høgde foreldre-barn, våre data

Estimat: $a = 43.3$ og $b = 0.77$



$$A = \bar{Y} - B\bar{x}$$

$$B = \frac{\sum_{i=1}^n (x_i - \bar{x}) Y_i}{\sum_{j=1}^n x_j^2 - n\bar{x}^2}$$

Minste kvadraters estimatorar

$$A = \bar{Y} - B\bar{x}$$

$$B = \frac{\sum_{i=1}^n (x_i - \bar{x}) Y_i}{\sum_{j=1}^n x_j^2 - n\bar{x}^2}$$

A og B er forventningsrette estimatorar: $E(A) = \alpha$ og $E(B) = \beta$ med varians

$$\text{Var}(A) = \sigma_A^2 = \frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n n(x_i - \bar{x})^2} \sigma_\epsilon^2$$

$$\text{Var}(B) = \sigma_B^2 = \frac{\sigma_\epsilon^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$A = \bar{Y} - B\bar{x}$$

$$B = \frac{\sum_{i=1}^n (x_i - \bar{x}) Y_i}{\sum_{j=1}^n x_j^2 - n\bar{x}^2}$$

A og B er forventningsrette estimatorar: $E(A) = \alpha$ og $E(B) = \beta$ med varians

$$\text{Var}(A) = \sigma_A^2 = \frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n n(x_i - \bar{x})^2} \sigma_\epsilon^2$$

$$\text{Var}(B) = \sigma_B^2 = \frac{\sigma_\epsilon^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Då Y_i ane er normalfordelte, og A og B er lineær kombinasjonar av Y_i ane, er

$$A \sim N(\alpha, \sigma_A^2)$$

$$B \sim N(\beta, \sigma_B^2)$$

Response: Det vi ønsker å predikere / forutsjå, Y
(stok.var.)

Forklaringsvariabel: Det vi kjenner/kontrollerer, x (skalar)

Enkel lineær regresjon: $Y = a + bx + \epsilon$, $\epsilon \sim N(0, \sigma_\epsilon^2)$